IEEE Transactions on Multimedia

A Novel Segmentation based Depth Map Up-sampling

Journal:	*IEEE Transactions on Multimedia*
Manuscript ID	MM-008180.R1
Suggested Category:	Regular Paper
Date Submitted by the Author:	21-Nov-2017
Complete List of Authors:	Qiao, Yiguo; Xidian University, School of Electronic Engineering Jiao, Licheng; Xidian University, School of Electronic Engineering Yang, Shuyuan; Xidian University, School of Electronic Engineering Hou, Biao; Xidian University, Key Lab of Intelligent Perception and Image Understanding of Ministry of Education of China
EDICS:	1-3DV 3-D Video Signal Processing < 1 SIGNAL PROCESSING FOR MULTIMEDIA APPLICATIONS, 2-PRES 3-D Processing and Presentation < 2 TECHNOLOGY COMPONENTS AND SYSTEM INTEGRATION, 1-IVGAS Image/Video/Graphics Analysis and Synthesis < 1 SIGNAL PROCESSING FOR MULTIMEDIA APPLICATIONS

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

Revised-TMM-A Novel Segmentation based Depth Map Up-sampling.rar

A Novel Segmentation based Depth Map Up-sampling

Yiguo Qiao, Licheng Jiao, Shuyuan Yang, and Biao Hou

***Abstract*—In this paper, we propose a novel color image segmentation guided depth map up-sampling method. In this method, the color image will be firstly segmented into a certain number of connected regions, then the target pixels will be interpolated by the seed pixels[1] regionally. In the segmentation phase, firstly the SLIC (simple linear iterative clustering) is introduced for generating superpixels, and the superpixels will be further divided into 8-connected regions. Secondly, the connected regions will be judged whether they are correct-clustered or not, the incorrect-clustered ones will be subdivided with an adaptive region-growing strategy. Thirdly, the regions that have no seed will be constantly merged into their nearest neighboring regions, until seed pixel can be found in each and every independent region. Lastly, adjacent regions that have quite small depth gaps will be united as one. The proposed color image segmentation strictly follows the guidance of the depth, and outputs credible connected regions which adheres to the depth boundary well. In the interpolation phase, the target pixels are interpolated with their surrounding seeds weighted by a joint trilateral filter (JTF). The proposed JTF is constructed by three terms, the color term, the distance term and the region term driven by the image segmentation result. Experimental results indicate that our method greatly reduces the depth bleeding and the depth confusion artifacts, and produces clear depth boundary in the up-sampled image. We also compare our method with the state-of-arts, comparisons verify the advantages of our method in both visual experience and quantitative evaluations.**

***Index Terms*—image segmentation, SLIC, region-growing, region merging, joint trilateral filtering**

I. Introduction

3D visual technologies have been widely used in multiple applications currently, such as free-viewpoint television (FTV), 3D games, 3D virtual video conference, human computer interaction and so on [1][2][3][4][5][6]. Accurate and high-resolution (HR) depth acquisition is one of the most crucial issues in lots of 3D visual technologies, which has attracted great attentions of many researchers.

For acquiring high quality HR depth maps, both passive based methods and active based methods have been put forward. The passive based methods usually obtain a depth map by calculating it with numbers of multi-view images [7]. However, both occlusions and structure-missing problems might be brought in during the calculating. The active based methods directly use some specific equipment to capture the depth maps, like the fusion camera system which is comprised of a color camera and a depth sensor [8][9]. A typical and widespread sensor is the Microsofts kinect [10], which can produce the depth map in real-time. However, the produced low resolution (LR) depth map could not cope with the demands. Therefore, LR depth up-sampling becomes critical important [11][12][13].

As the depth up-sampling is under the guidance of the registered HR color image, it is totally different from the image super-resolution system [14][15][16]. In theory, up-sampling rate can be relatively high while guaranteeing high-quality results.

A. Related Work

Lots of depth up-sampling methods have been proposed until now. Kopf et al. proposed the joint bilateral up-sampling (JBU) method [17] in 2007. By using a local joint bilateral filtering, the authors interpolate the LR depth map into HR. The filter is the product of two Gaussian kernels, which are used to describe the spatial distance and the color difference, respectively. As a local algorithm and only two factors are taken into consideration, this method produces unsatisfactory up-sampling result. Especially in the case of comparatively large up-sampling rate, depth missing and depth confusion will be produced.

Y.S. Ho et al. proposed the joint bilateral plus local minimum (JBLM) filtering based method [18] and the Markov random field (MRF) model based method [19] in 2013 and 2014, respectively. The JBLM method is on the basis of JBU. The authors divide the joint bilateral up-sampled result into continuous regions and discontinuous regions, and adopt the LM filtering in those depth discontinuous regions. This method somewhat reduces the depth confusion artifacts in the boundary regions. In the MRF model based method, the authors construct an energy function in MRF for solving the up-sampling problem. This energy function is optimized via belief propagation [20]. However, over-smoothing might be produced during the optimization.

M. Liu et al. proposed the classical joint geodesic up-sampling (JGU) method [21] in 2013. In this method, the authors represent the definition of the geodesic distance firstly. Then the geodesic distances between the targets and the seeds are calculated by a recursive algorithm. Lastly, the targets are interpolated with the nearest several seeds according to the calculated distance. Though this global algorithm greatly solves the depth missing problem, the depth bleeding artifacts still can be found.

D. Ferstl et al. proposed the anisotropic total generalized variation (TGV) based up-sampling method [22] in 2013. In their work, the up-sampling is formulated as a global energy

[1]The seed pixels are directly from the low resolution depth maps, i.e. the ones that have depth values. The targets are those without depth and to be interpolated.

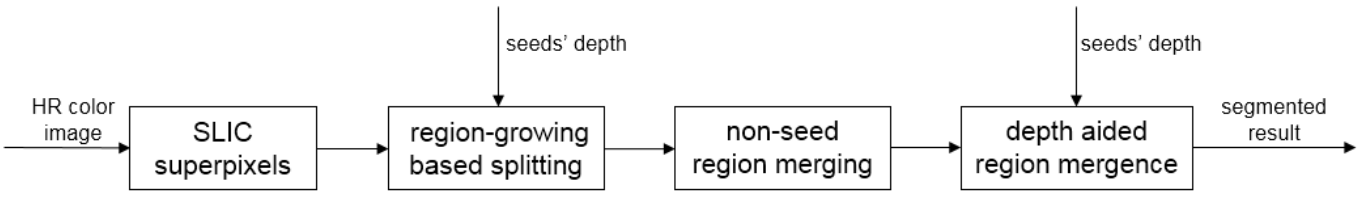


Fig. 1. Flow chart of the proposed color image segmentation.

optimization problem. The energy function is a combination of the data term and the TGV term that includes an anisotropic diffusion tensor. The authors solve this optimization problem by using a first order primal-dual algorithm. Since the tensor not only weights the depth gradient but also orients the gradient direction, this method preserves the depth boundary well. However, over-smoothing comes into being across the homogeneous regions.

J. Yang et al. proposed the adaptive autoregressive model based method [23] in 2014. The authors formulate the depth recovery task as a minimization of the AR prediction errors subject to measurement consistency. The proposed method works well but the time and memory costs are considerably large.

J. Kim et al. proposed the joint adaptive bilateral depth up-sampling (JABDU) method [24] in 2014. The authors firstly adjust a color filter and a depth filter based on an adaptive smoothing parameter and a control parameter. They the targets are interpolated with their neighbors weighted by the above adjusted filters. Depth confusion could not be completely avoided, especially when under a relatively large up-sampling rate.

K. Lo et al. proposed the joint trilateral filter based up-sampling [25] in 2017, the proposed method is on the basis of the bi-cubic interpolation result. The authors refine the intermediate result by using a joint trilateral filter in an outside-inward refining order regularized by the detected color/depth edges. This method preserves the edges very well when the up-sampling rate no larger than 4, but not so satisfying as the up-sampling rate increases.

B. Overview of the Proposed Method

The proposed method aims to generate the HR depth map with clear boundary even in the case of large up-sampling rate. The proposed method can be decomposed into two phases: the color image segmentation and the depth map interpolation.

The flow chart of the proposed color image segmentation is shown in Fig.1, which mainly consists of 4 steps:

1) Firstly, the SLIC superpixels is introduced as a rough segmentation, and the connected regions of each superpixel are obtained;
2) Secondly, with the aid of the given splitting threshold, the obtained connected regions are judged whether they are correct-clustered or not. Then those incorrect-clustered regions will be split into several subregions based on a depth guided region-growing strategy;
3) Thirdly, we merge the regions that have no seed into their nearest neighbors through a loop. Ensure that seed pixel can be found in every independent region;
4) Fourthly, the adjacent regions, the depth gaps between which are smaller than the merging threshold, are combined into one. Loop this depth aided region mergence algorithm until there is no region can be combined.

The segmentation result based on the above proposed method is shown in Fig.2. Then the targets will be interpolated with the seeds by using the proposed JTF, which is weighted by three terms. They are the color term, the distance term and the region term derived from the segmentation result. Benefits from the co-work of the three terms, the proposed JTF owns good robustness.

C. Contributions

- We adopt the SLIC as a pre-segmentation, which outputs compact and uniformly distributed superpixels.
- We introduce a depth-guided adaptive region-growing to split the incorrect-clustered regions, pixels in these regions may have similar colors but distinct depth values. We further adopt a depth aided region mergence to combine the regions that have similar depths. With this depth guided segmentation first and recombination second strategy, useful depth boundary is preserved well.
- We put forward a JTF based interpolation with consideration of three factors mentioned above. Target pixels are interpolated robustly with the proposed JTF.

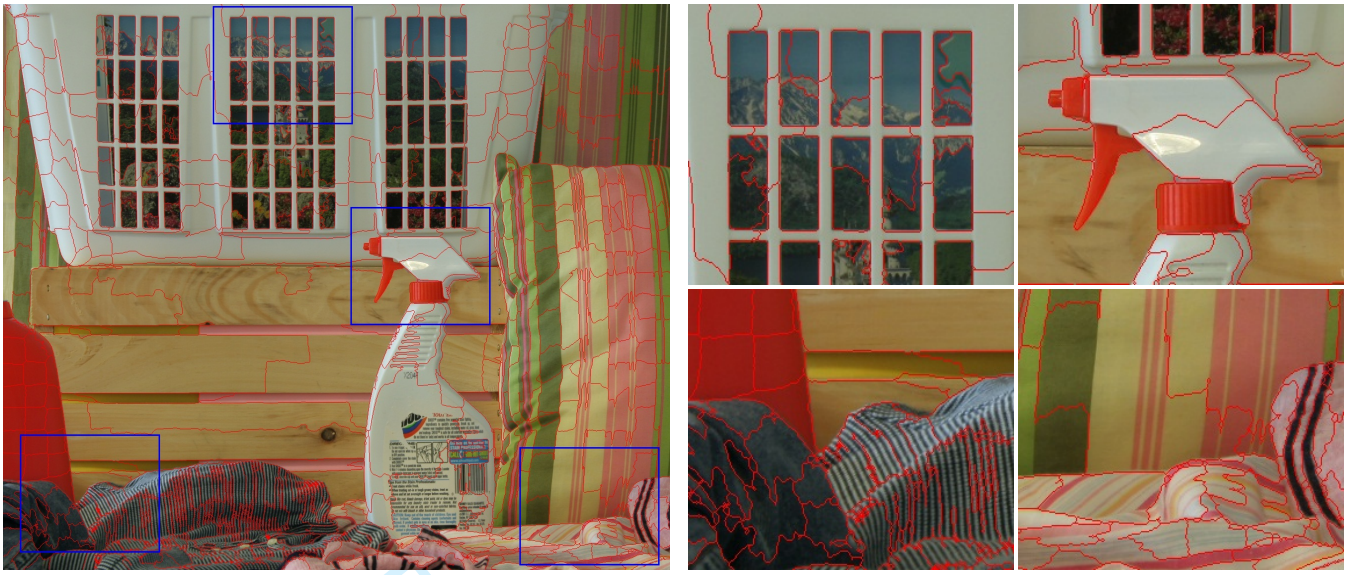
The rest of the paper is organized as follows. Section II gives a briefly introduction of the depth guided color image segmentation. Section III presents the proposed JTF. Experimental results and parameter analysis are provided in Section IV. We conclude this paper in Section V.

II. Depth Guided Color Image Segmentation

In this section, the HR color image I will be segmented into a certain number of connected regions with the guidance of the corresponding LR depth map D_L. 4 steps are included in this depth guided color image segmentation, a) SLIC superpixels, b) region-growing based splitting, c) non-seed region merging, d) depth based region mergence.

A. SLIC Superpixels

In this step, the SLIC method is introduced to segment the color image into compact superpixels. Actually, all graph-cut based or clustering based over-segmentation methods can



(a) (b)

Fig. 2. The final segmentation result. (a): the segmented color image. (b): the enlarged details.

be embedded in the proposed framework, such as Ncuts[26], Turbopixels[27], MSLIC [28] and so on. SLIC is a reworking of the k-means clustering [29], it is selected in considering of both the computation speed and the edge-preserving performance [30].

Firstly, we convert the HR color image I into the CIELAB color space [31]. The CIELAB mode is selected because it is more closely to humans visual physiology, and it is only relevant to the color property rather than the equipment. Then we combine the 3 color channels $[l, a, b]$ with the pixel coordinate $[x, y]$ to form a 5-D vector space $[l, a, b, x, y]$. The similarity of any two pixels can be measured by calculating the Euclidean distance of their vectors. The greater the distance, the weaker the similarity.

Then with a given initial number of superpixels k, we determine the k initial superpixel centers and set the side length of each superpixel $S = \sqrt{N/k}$, where N denotes the total number of pixels in the input color image. For avoiding the superpixel centers fell on noise points or contour boundaries, we reselect them within their $3*3$ neighborhoods. The one with the smallest gradient among the candidates will be chosen.

Next we calculate the geodesic distances between each center pixel and its neighbors located in a $2S * 2S$ search region. The geodesic distance D between a center pixel i and one of its neighbours j can be defined as Eq.(1) shows,

$$D = \sqrt{(\frac{D_c}{s})^2 + (\frac{D_s}{S})^2} \tag{1}$$

where D_c and D_s can be calculated by Eq.(2) $\sim$ Eq.(3), they stand for the color distance and the spatial distance, respectively. Since the intensities of l, a and b channels are limited but does not the image size, thus if the image size is relatively large, an excessive proportion of the spatial distance will be brought in the calculation of the geodesic distance. Thus a scale constant s is introduced to adjust and control the proportion of the spatial distance.

$$D_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \tag{2}$$

$$D_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{3}$$

Based on the calculated geodesic distances, we label each pixel with its nearest superpixel center. When all pixels are labeled, we update the k superpixel centers by averaging the vectors of all pixels in each superpixel. Loop this pixel clustering process, including the distance calculating, the pixel labeling and the center updating, until converges.

As known, some un-connected regions that have quite similar colors, may be clustered into one superpixel. However, the depths of those regions might be quite different since disconnected. Due to this, we separate the superpixels into 8-connected regions. And for accelerating the program, we unit the isolated points into their neighboring regions. Until now, we obtain the rough segmented result Ψ, which contains a number of M connected regions and can be expressed as $\Psi = \{\Psi_m \mid 1 \leq m \leq M\}$.

B. Region-growing based Splitting

Firstly, we project the LR depth map D_L onto the HR grid for generating the intermediate HR depth map d, which is the same size as the HR color image. Then based on this intermediate HR depth map, we separate the above connected regions among Ψ into 3 categories. The first category are the non-seed regions, i.e. the regions that have no seed pixel. The rest, i.e the regions have seed pixels within themselves, can be classified into the other two categories, the correct-clustered regions and the incorrect-clustered regions. The region-growing is used to split those incorrect-clustered regions into correct-clustered subregions.

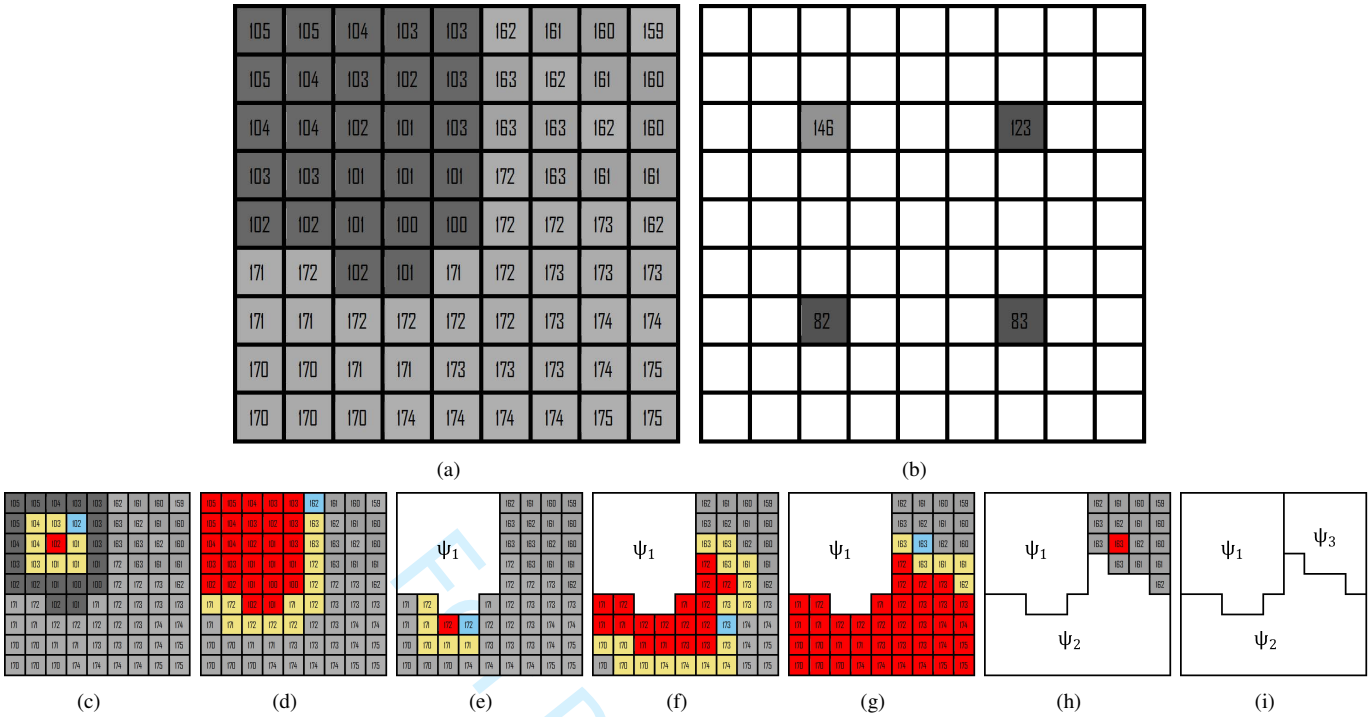


Fig. 3. Region-growing based Splitting. Pixels labeled in red are already segmented, labeled in yellow are the surrounding pixels of the segmented subregion, labeled in blue are the next growing points. (a) Uncorrect-clustered color region. (b) The corresponding depth region. (c) Initial point selection of the first subregion. (d) Judgment of growing terminal condition I (being met). (e) Termination of the growth of the first subregion and initial point selection of the second subregion. (f) Judgment of growing terminal condition II (when not being met). (g) Judgment of growing terminal condition II (being met). (h) Termination of the splitting of the region. (i) The generated subregions.

Region-growing is actually a process of polymerizing independent pixels or small regions into larger regions [32][33], based on a predefined criteria. As known, the performance of the region-growing is mainly depends on three factors, the selection of the initial point, the criterion of the growth and the terminal condition. In our proposed depth guided adaptive region-growing method, the initial points is selected adaptively with the guidance of the depth, and the color threshold which controls the termination of the growth is also self-adjusted according to various regions.

1) Region classification: Firstly, we divide the seeds from the targets in a current region Ψ_m as Eq.(4) shows,

$$S_{\Psi_m} = \{p \mid p \in \Psi_m \& d(p) > 0\}. \tag{4}$$

where S_{Ψ_m} denotes the seeds set in Ψ_m; p denotes the pixels in the HR image; $d(p)$ is the depth of p. If S_{Ψ_m} is non-empty, we sort the depth of the seeds in S_{Ψ_m} in ascending order as eq.(5) shows,

$$d_{\Psi_m}{}' = F\{d(S_{\Psi_m})\}. \tag{5}$$

where F is a function of ascending order; $d_{\Psi_m}{}'$ represents the sorted depth values.

Then we judge which category does the current region Ψ_m belongs to through Eq.(6),

$$C(\Psi_m) = \begin{cases} -1, & S_{\Psi_m} = \emptyset \\ 0, & S_{\Psi_m} \neq \emptyset \ \& \ \forall i,\ d_{\Psi_m}{}'(i+1) - d_{\Psi_m}{}'(i) \leq th_s \\ 1, & else \end{cases} \tag{6}$$

where $i \in (0, q)$, q denotes the length of $d_{\Psi_m}{}'$; th_s is a given splitting threshold. More specifically, if the seeds set of the current region Ψ_m is empty, i.e $C(\Psi_m) = -1$, Ψ_m belongs to non-seed regions. Otherwise, if all of the depth gaps between two arbitrary seeds are no larger than the threshold th_s, i.e. $C(\Psi_m) = 0$, the current region Ψ_m will be classified into correct-clustered regions, or once there exists a depth gap larger than the threshold th_s, i.e. $C(\Psi_m) = 1$, Ψ_m will be regarded as an incorrect-clustered region.

Once the current region Ψ_m is incorrect-clustered, i.e. $C(\Psi_m) = 1$, we isolate subregions from it one after another until the remaining Ψ_m is correct-clustered. Fig.3 demonstrate the splitting of an incorrect-clustered region in detail (for simplicity, we make the region rectangular). Fig.3(a) shows the to be split color region in LAB mode, the corresponding depth region is shown in Fig.3(b).

2) Initial point selection: We choose the initial growing point of a current subregion ψ among the seed pixels in Ψ_m. That is because whether the remaining Ψ_m is correct-clustered or not depends on the depths of the remaining seeds, thus growing from the seed pixels can significantly reduce the splitting times. Among the seeds, the one with the maximum absolute deviation (as the pixel labeled in red in Fig.3(c)) will be selected as the initial growing point since it can accelerate the convergence of the splitting. Eq.(7) formulates the selection of the initial growing point i' in the current subregion ψ, and it will be labeled as the first segmented point

in ψ.

$$i' = \arg\max_{i} \mid d(S_{\Psi_m}(i)) - \frac{1}{m}\sum_{j=1}^{m} d(S_{\Psi_m}(j)) \mid, \quad s.t.\ 1 \le i \le m. \tag{7}$$

Next we press the 8 neighbors of i' into a linked list L, i.e. $L = N_{i'}$, where $N_{i'}$ denotes the 8 neighbors of i'. The linked list L is used to save the surrounding pixels, i.e. the candidates of the next growing point, of a current subregion ψ.

3) Growing criteria: The next growing point l' among L can be selected through a growing criteria shown in Eq.(8),

$$l' = \arg\min_{l} \mid \frac{\sum_{i\in\psi} Lab(i)}{\sum_{i\in\psi} 1} - Lab(l) \mid, \quad s.t.\ l \in L. \tag{8}$$

where $Lab(\psi)$ is the sum of l, a, b channels of all pixels in ψ. It is easy to understand that we select the pixel whose color is the nearest to the mean color of the current subregion ψ as the next growing point. Pixels labeled in blue in Fig.3 represent the next growing points in each step.

Note that, l' is only a to be merged pixel of ψ until now, it does not means that l' must will be merged into ψ. Whether it can be merged or not depends on the following growing terminal conditions.

4) Growing terminal conditions: The growing terminal conditions depend on the colors and the depths are shown in Eq.(9) and Eq.(10), respectively.

$$ConditionI: \mid \frac{\sum_{i\in\psi} Lab(i)}{\sum_{i\in\psi} 1} - Lab(l') \mid > th_c(\Psi_m). \tag{9}$$

$$ConditionII:\ d(l') \neq 0\ \&\&\ \min_{j\in S_\psi} \mid d(j) - d(l') \mid > th_s. \tag{10}$$

The first terminal condition is used to limit the color of the to be merged pixel l'. Once the color difference between l' and the mean of the current subregion ψ is larger than the region-adaptive color threshold $th_c(\Psi_m)$, the growing terminates. Fig.3(d)) illustrates this situation visually.

Moreover, too small color threshold will lead over-segmentation, and too large makes the splitting meaningless. Empirically, we set the color threshold equals the mean absolute deviation (MAD) of the splitting region Ψ_m as Eq.(11) shows,

$$th_c(\Psi_m) = \frac{\sum_{k\in\Psi_m} \mid Lab(k) - Mean(Lab(\Psi_m)) \mid}{\sum_{k\in\Psi_m} 1} \tag{11}$$

where $Mean(i)$ denotes the mean value of i.

The second terminal condition is set against the depths, and it makes sense if and only if when the growing point l' goes into a seed pixel, since only the seeds have depth values. S_ψ in Eq.(10) denotes the seeds set in the current subregion ψ, which can reference in Eq.(4). Once the minimum depth difference between one of the seeds in ψ and the growing point l' is larger than the previous splitting threshold th_s, the growing of the current subregion ψ terminates. Fig.3(f) presents the situation that the minimum depth difference is smaller than the splitting threshold, the opposite situation is presented in Fig.3(g).

If none of the conditions above is met, the growing of ψ continues. In this case, we merge the growing point l' into ψ and remove it from the linked list L. Then, among the 8 neighbors of l', those who are not included in L yet will be pressed into.

Repeat from Sec.II-B3 until either or both of the above two conditions are met. And until then, the growing of the current subregion ψ terminates. Then we add ψ as a newly independent subset into the region set Ψ and update the current region Ψ_m by dividing ψ from it. Fig.3(e) and Fig.3(h) present the generation of the newly subregions under the above two different situations.

Repeat from Sec.II-B1 until Ψ_m is correct-clustered, i.e. $C(\Psi_m) = 0$, then we terminate the splitting of region Ψ_m (as shown in Fig.3(i)) and turn to a next region.

Algorithm1 summarizes the procedure of the proposed region-growing based splitting. The algorithm strictly follows the guidance of the depth and further leads a credible splitting result that adheres to the image boundary well. However, holes led by noises and inhomogeneous intensities will be produced during the region growing, it makes some regions that have a same label disconnected. Thus we separate those disconnected regions into 8-connected subregions and get the final region set Ψ after the region-growing based splitting.

In the current fine segmentation result, lots of adjacent regions that have similar depths are independent each other, it is actually overdone for depth map up-sampling. What is more, tiny regions might have no seed pixel to interpolate the targets among them regionally. For these above two reasons, we introduce the region merging in Sec.II-C and Sec.II-D.

C. Non-seed Region Merging

Suppose there are N independent regions in the current region set Ψ after the above splitting process. Only two categories are remained among the N regions, the correct-clustered regions and the non-seed regions. According to different types of regions, we separate the proposed region merging into two phases [34] . In the first phase, we merge the non-seed regions into their neighboring correct-clustered regions with the guidance of color. And in the second phase we merge the adjacent regions that have similar depths into a whole with the guidance of both color and depth. We present the first non-seed region merging in this section and the second depth aided region mergence in the next section.

Suppose a current region Ψ_n is an non-seed region judged by Eq.(6), and $\{\Psi_{n_1}, \Psi_{n_2}, \cdots, \Psi_{n_U}\}$ are its U neighbours. We choose the recipient $\Psi_{n_{u'}}$ as Eq.(12) shows,

$$u' = \arg\min_{u} \mid Mean(Lab(\Psi_{n_u})) - Mean(Lab(\Psi_n)) \mid, \quad s.t. \quad 1 \le u \le U\ \&\&S_{\Psi_{n_u}} \neq \emptyset. \tag{12}$$

where $S_{\Psi_{n_u}}$ denotes the seeds set in Ψ_{n_u}, $S_{\Psi_{n_u}} \neq \emptyset$ limits that the recipient $\Psi_{n_{u'}}$ must be a correct-clustered region. Then we

Algorithm 1 Region-growing based splitting.

Input:
The intermediate HR depth map d;
The HR color image Lab;
The region set $\Psi = \{\Psi_1, \Psi_2, \cdots, \Psi_M\}$;
The splitting threshold th_s;
Output:
The region set Ψ after splitting;

for each $m \in [1, M]$ **do**
  Initialize $C(\Psi_m) = 1$;
  repeat
    Get the seeds in Ψ_m as in Eq.(4) and sort their depth in ascending order as in Eq.(5);
    Calculate $C(\Psi_m)$ in Eq.(6);
    if $C(\Psi_m) = 1$ **then**
      Select the initial point i' using Eq.(7);
      Initialize the linked list $L = N_{i'}$;
      Initialize $Case^2 = 1$;
      repeat
        Find the next growing point l' in Eq.(8);
        Calculate $Case$ through Eq.(9) $\sim$ Eq.(11);
        if $Case = 1$ **then**
          $\psi \Leftarrow \psi \bigcap l'$;
          $L \Leftarrow L \setminus l'$;
          $L \Leftarrow L \bigcap N_{l'}$;
        end if
      until $Case \neq 1$
      $\Psi \Leftarrow \{\Psi, \psi\}$;
      $\Psi_m \Leftarrow \Psi_m \setminus \psi$;
    end if
  until $C(\Psi_m) \neq 1$
end for

merge the non-seed region Ψ_n into $\Psi_{n_{u'}}$, and remove Ψ_n from the region set Ψ.

From $n = 1$ to N, we conduct the non-seed region merging on all non-seed regions in Ψ. After that, seed pixel can be found in each and every independent region.

D. Depth aided Region Mergence

Suppose $\Psi = \{\Psi_1, \Psi_2, \cdots, \Psi_T\}$ is the region set at present. Then from Ψ_1 to Ψ_T, we judge whether a current region Ψ_t satisfies the merging criteria, and merge the satisfied one into its nearest neighbor.

Let $\{\Psi_{t_1}, \Psi_{t_2}, \cdots, \Psi_{t_V}\}$ be the V neighboring regions of the current region Ψ_t. Then we calculate the maximum depth difference between Ψ_t and each of the V neighbors as Eq.(13) shows,

$$dep_v = \max(d(S_{\Psi_t}; S_{\Psi_{t_v}})) - \min(d(S_{\Psi_t}; S_{\Psi_{t_v}})),\ 1 \leq v \leq V. \tag{13}$$

where S_{Ψ_t} and $S_{\Psi_{t_v}}$ denote the sets of the seeds in region Ψ_t and Ψ_{t_v}, respectively; $\max(d(S_{\Psi_t}; S_{\Psi_{t_v}}))$ and $\min(d(S_{\Psi_t}; S_{\Psi_{t_v}}))$ denote the maximum depth and the minimum depth of the seeds in the united region $(\Psi_t; \Psi_{t_v})$, respectively; dep_v provides the maximum depth difference of the seeds in the current region Ψ_t and its neighbor Ψ_{t_v}.

Algorithm 2 Depth aided region mergence.

Input:
The intermediate HR depth map d;
The HR color image Lab;
The current region set $\Psi = \{\Psi_1, \Psi_2, \cdots, \Psi_T\}$;
The merging threshold th_m;
Output:
The final region set Ψ;

repeat
  for $t = 1$ to T **do**
    Find the neighbours of Ψ_t, suppose they are $\Psi_{t_1}, \Psi_{t_2}, \cdots, \Psi_{t_V}$;
    for each $v \in [1, V]$ **do**
      Calculate the maximum depth difference dep_v between Ψ_t and Ψ_{t_v} as in Eq.(13);
    end for
    Find the neighbour regions $\Psi_{t_{V'}}$ that have similar depths with Ψ_t, i.e search out V' in Eq.(14);
    if V' is empty **then**
      $t \Leftarrow t + 1$;
      Continue;
    else
      Find the nearest neighbour $\Psi_{t_{v''}}$ through Eq.(15);
      $\Psi_{t_{v''}} \Leftarrow \Psi_t \bigcap \Psi_{t_{v''}}$;
      Removing Ψ_t from Ψ;
      $T \Leftarrow T - 1$;
    end if
  end for
until The region number T converges

Through the merging criteria shown in Eq.(14), we judge whether Ψ_t can be merged into one of its neighbors,

$$V' = \{v \mid dep_v \leq th_m\},\ 1 \leq v \leq V. \tag{14}$$

where V' denotes a set of region labels. It can be summarized that all regions in the region set $\{\Psi_{t_{V'}}\}$ have similar depths with the current region Ψ_t. If V' is empty, it means that Ψ_t has large depth gaps with all its neighbours. Thus it can be merged into none of its neighbors, and we will skip to the next region Ψ_{t+1}. Otherwise, Ψ_t can be merged into a specific region in $\{\Psi_{t_{V'}}\}$. We select this specific region with the help of the color as Eq.(15) shows,

$$v'' = \arg\min_{v'} \mid Mean(Lab(\Psi_t)) - Mean(Lab(\Psi_{t_{v'}})) \mid,$$
$$s.t. \qquad v' \in V'. \tag{15}$$

where v'' denotes the label of the specific region, the obtained region $\Psi_{t_{v''}}$ is regarded as the nearest neighbor of the current region Ψ_t. Then we merge Ψ_t into the specific region $\Psi_{t_{v''}}$ and update the region set Ψ by removing Ψ_t from it.

Loop this mergence until the region number in Ψ converges, and at that time we get the final segmentation result. The procedure of the depth aided region mergence algorithm is summarized by Algorithm2.

Performance of the proposed depth guided image segmentation is shown in Fig.4, from which we can see that

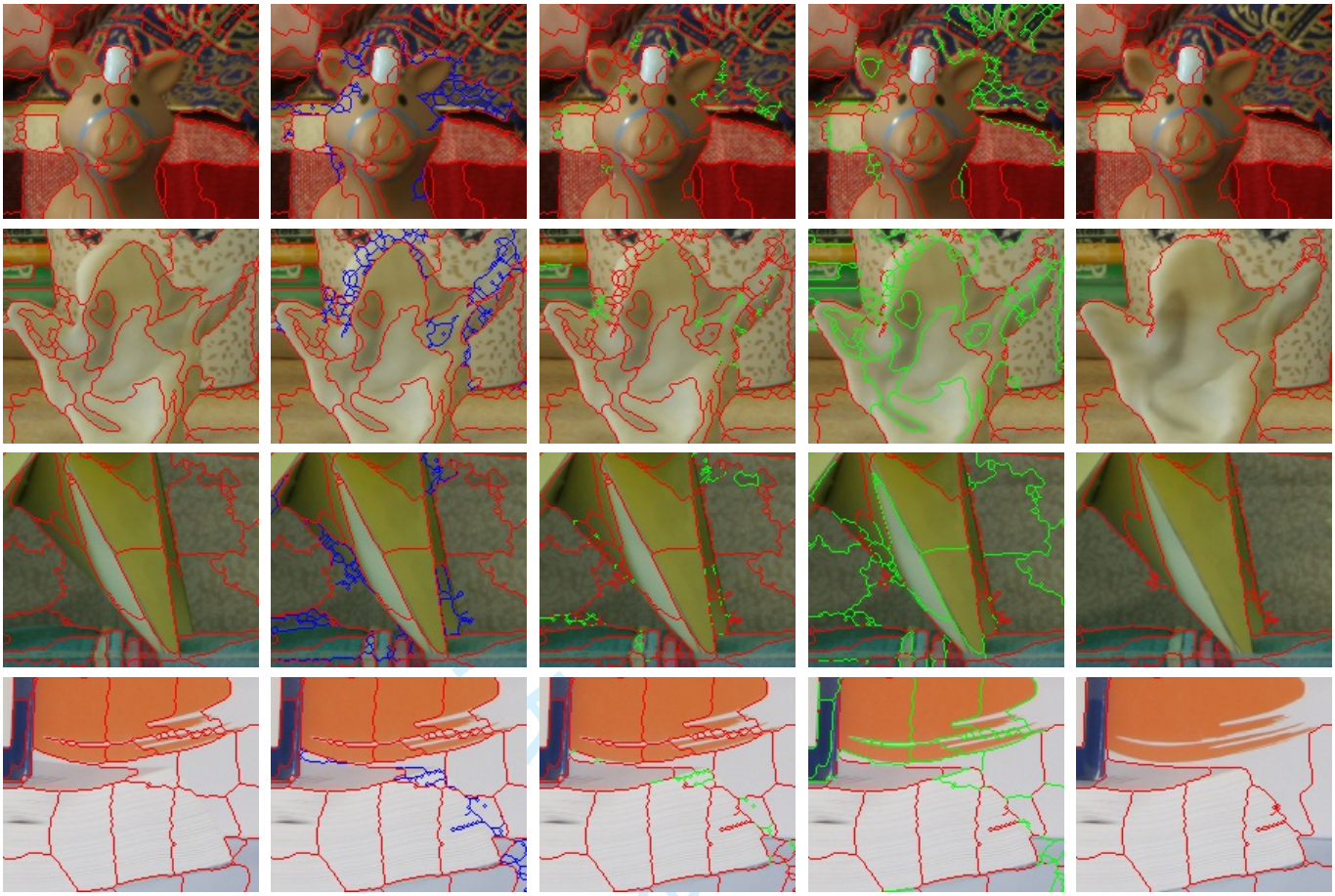


Fig. 4. Performance of the proposed segmentation method. (Blue lines indicate the new adding boundaries during the region splitting process, green lines indicate the missing boundaries during the region merging process.) (a): segmented results after the SLIC superpixels; (b): segmented results after the region-growing based splitting; (c): segmented results after the non-seed region merging; (d): segmented results after the depth based region mergence; (e): the final segmented results.

the proposed method separates the regions have large depth gaps and combines the regions with close depth values. The segmented result directly determines the region term in the following JTF.

III. JTF Based Interpolation

The proposed JTF based depth interpolation is inspired by the conventional JBF. Three weighting terms, the color term, the distance term and the region term are involved in the proposed JTF [35][36].

A target pixel t is interpolated through the JTF as Eq.(16) shows,

$$d(t) = \frac{\sum_{s \in N_t} d(s) \cdot W(s,t)}{\sum_{s \in N_t} W(s,t)} \tag{16}$$

where N_t is a rectangular window centered at t, for simplicity, its search range can be double or triple of the up-sampling rate; $W(s,t)$ denotes the weighting coefficient and can be defined as in Eq.(17),

$$W(s,t) = w_c(s,t) \cdot w_d(s,t) \cdot w_r(s,t) \tag{17}$$

where w_c, w_d and w_r represent the color term, the distance term and the region term, respectively.

Among the three terms, w_c and w_d are simulated by two Gaussian functions as shown in Eq.(18) and Eq.(19),

$$w_c(s,t) = G_{\sigma_c}(\mid Lab(s) - Lab(t) \mid) \tag{18}$$

$$w_d(s,t) = G_{\sigma_d}(\parallel s - t \parallel) \tag{19}$$

where σ_c and σ_d denote the variances. The color term w_c reflects the color difference between two pixels, and outputs higher weights when the two pixels have similar intensities. The distance term w_d reflects the spatial distance between two pixels, and provides larger proportion when the two pixels are spatially close to each other.

The region term w_r is derived by the segmented result and can be described as in Eq.(20),

$$w_r(s,t) = \begin{cases} 1, & s \in \Psi_p \\ \gamma, & s \in \Psi_p{}' \\ 0, & else \end{cases} \tag{20}$$

where Ψ_p denotes the region that the target pixel t located in, $\Psi_p{}'$ denotes the neighboring regions of Ψ_p; γ is a parameter

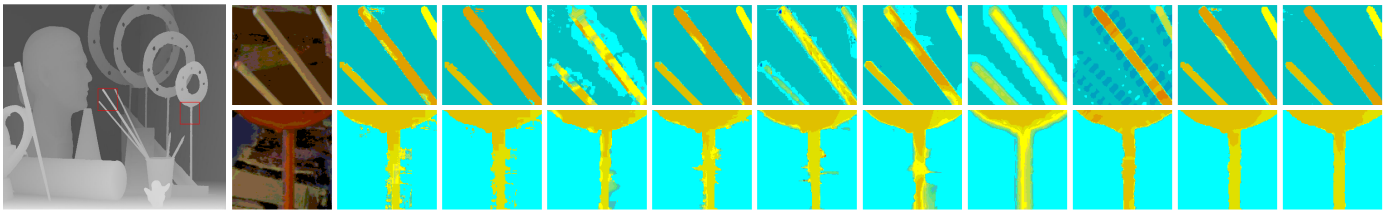


Fig. 5. 8 times up-sampling results of *Art*. (a) Ground truth depth map, (b) color regions, and up-sampling results of (c) JBU, (d) JBLM, (e) Mrf, (f) JGU, (g) TGV, (h) AR Model, (i) JABDU, (j) JTF, (k) the proposed method, and (l) associated ground truth.

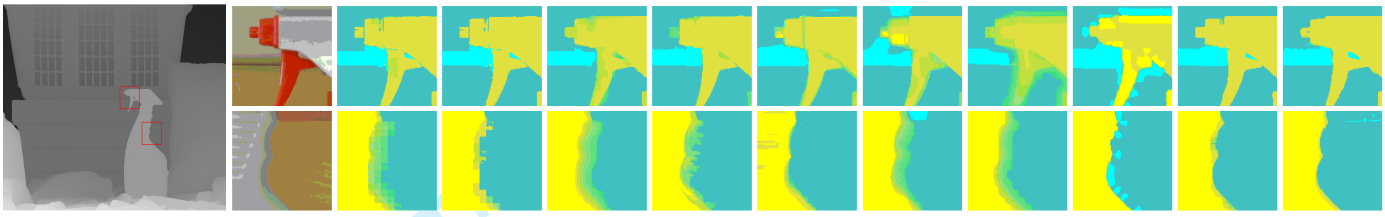


Fig. 6. 8 times up-sampling results of *Laundry*. (a) Ground truth depth map, (b) color regions, and up-sampling results of (c) JBU, (d) JBLM, (e) Mrf, (f) JGU, (g) TGV, (h) AR Model, (i) JABDU, (j) JTF, (k) the proposed method, and (l) associated ground truth.

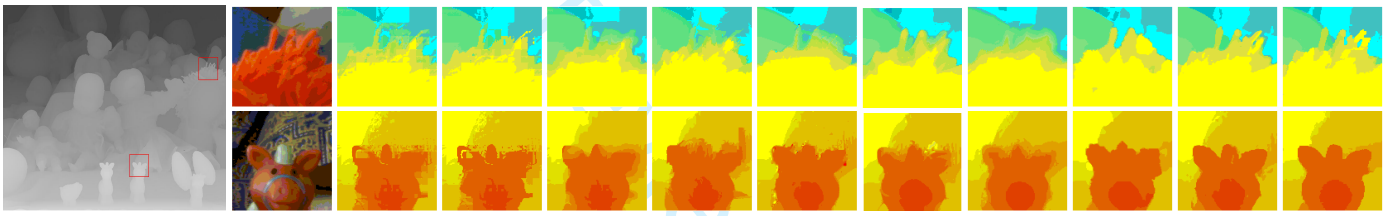


Fig. 7. 8 times up-sampling results of *Dolls*. (a) Ground truth depth map, (b) color regions, and up-sampling results of (c) JBU, (d) JBLM, (e) Mrf, (f) JGU, (g) TGV, (h) AR Model, (i) JABDU, (j) JTF, (k) the proposed method, and (l) associated ground truth.

in $(0, 1)$. It indicates that a high weight up to 1 will be returned to w_r when the seed pixel is in the same region with the target, and being lower when the seed is in one of the neighbouring regions of the target. Otherwise if the seed is in some other distant regions, the region weight w_r drops to 0. From the above, only the seeds in the same region and the surrounding regions of the target will be weighted for the interpolation. Furthermore, the region term also makes it possible for the search window to have a relatively large search range. Therefore, homogeneous seeds even if far away from the targets, can be searched out and utilized for interpolating.

Through the proposed JTF, we generate the up-sampled depth map with clear boundary even if under a relatively large up-sampling rate. Besides, for accelerating the program, the interpolation would better be conducted region by region.

IV. Experimental Results

A. Visual and Quantitative Evaluations

We conduct the experiments on a PC with Core Duo 3.20GHz CPU and 4.0G RAM by using Matlab. Six data sets from the Middleburys benchmark [37], including *Art*, *Laundry*, *Dolls*, *Book*, *Moebius*, and *Reindeer*, are used for evaluation.

Two types of degradations are simulated in our experiments: unharmed depth down-sampling and Kinect-like depth down-sampling [23]. In the first type, no depth missing or noising exists in the depth maps. Thus, we recover the depth maps in the database before the experiments [38]. In the second type, structural missing and random missing can be found in the depth maps. Up-sampling results according to these two different degradation types are shown as below.

We use the bad pixel rate (BP) and the root-mean-square error (MSE) to assess the up-sampling performance [25]. Bad pixels are those whose depth value in the up-sampled image deviates from the ground truth by more than one disparity. The BP and the MSE statistics reflect the quantities and the degree of the errors, respectively. These two metrics are applied in the disocclusion regions (disocc.), the entire image (all) and the depth discontinuous regions (disc.), respectively.

Disocclusions denote the structural-missing regions and the random-missing regions, i.e. the regions that have no depth information. Evaluation of the disocclusion regions verifies the performance of the algorithm on those depth-missing areas.

Depth discontinuous regions are detected with a $5*5$ sliding window, if the maximum depth difference of the pixels in the current window is larger than a given depth threshold, the pixel that the current window located on will be regarded as a depth

TABLE I
QUANTITATIVE UP-SAMPLING RESULTS FROM UNHARMED DEPTH DOWN-SAMPLING WITH UP-SAMPLING FACTOR OF 8 (IN BP). FOR EACH IMAGE, THE BEST PERFORMANCE IS SHOWN IN BOLD.

Methods	Laundry			Dolls			Art			Books			Moebius			Reindeer		
	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.
JBU	6.00	6.30	26.29	9.66	10.07	38.60	11.75	12.23	40.99	7.45	7.91	38.89	8.65	9.22	44.29	7.96	8.37	37.58
JBLM	4.56	4.84	20.28	8.56	8.97	32.57	7.14	7.62	**21.16**	6.60	7.06	32.76	6.73	7.28	34.43	6.45	6.83	25.09
MRF	7.29	7.58	34.85	8.96	9.35	41.73	12.76	13.24	48.66	8.01	8.50	48.10	9.03	9.50	52.10	8.04	8.61	46.64
JGU	5.71	6.00	28.33	8.53	9.02	40.33	6.97	7.51	27.33	7.46	7.99	41.86	7.35	7.89	42.31	5.24	55.86	36.85
TGV	7.37	7.65	32.04	6.20	6.65	38.10	6.31	6.84	43.33	3.45	4.05	32.94	4.94	5.60	36.67	4.36	4.95	40.56
AR	7.03	7.36	33.91	6.69	6.14	36.32	6.51	**6.60**	40.63	**3.31**	**3.93**	31.35	4.68	5.31	33.97	4.15	4.33	35.15
JAB	7.62	7.90	40.10	7.18	7.59	47.26	13.29	13.75	65.04	6.14	6.64	57.10	7.55	8.10	53.95	7.27	7.74	63.91
JTF	7.94	8.26	45.29	8.57	8.71	45.76	6.92	7.42	47.80	6.60	7.12	60.14	6.89	7.41	48.40	**2.79**	**3.13**	33.69
Ours	**2.63**	**2.91**	**16.48**	**4.74**	**5.22**	**25.09**	**6.27**	6.79	24.65	3.92	4.37	**30.27**	**4.66**	**5.14**	**27.74**	3.67	3.99	**24.09**

TABLE II
QUANTITATIVE UP-SAMPLING RESULTS FROM UNHARMED DEPTH DOWN-SAMPLING WITH UP-SAMPLING FACTOR OF 8 (IN MSE). FOR EACH IMAGE, THE BEST PERFORMANCE IS SHOWN IN BOLD.

Methods	Laundry			Dolls			Art			Books			Moebius			Reindeer		
	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.
JBU	3.32	3.57	8.79	2.28	2.39	6.14	6.61	6.80	17.20	2.87	3.11	10.11	2.71	3.03	7.91	3.49	3.98	13.07
JBLM	3.58	3.85	9.61	2.41	2.55	6.66	7.22	7.44	19.04	3.17	3.40	11.13	2.14	3.29	8.75	3.71	4.23	14.09
MRF	**2.50**	**2.68**	**6.79**	1.72	1.82	4.74	6.25	6.44	17.13	**2.21**	**2.41**	**8.02**	2.09	2.81	7.03	3.18	4.24	12.65
JGU	2.98	3.29	8.43	2.21	2.37	6.23	5.44	5.73	15.37	2.68	2.89	9.35	2.42	2.61	7.26	3.30	3.65	12.53
TGV	3.08	3.48	8.53	1.85	1.96	5.25	5.97	6.20	16.82	2.19	2.48	8.82	2.25	2.59	7.25	4.30	4.59	13.31
AR	2.77	3.10	8.11	1.80	1.90	5.15	5.64	5.82	15.38	2.52	2.78	9.37	2.11	2.38	6.66	3.11	3.54	12.12
JAB	2.90	3.06	8.20	1.78	1.86	5.22	6.14	6.27	17.18	2.66	2.86	10.51	**2.06**	**2.21**	**6.49**	3.85	4.17	14.67
JTF	4.00	4.22	10.63	2.31	2.38	5.71	6.65	7.02	18.32	2.64	2.92	10.60	2.45	2.72	8.11	4.38	4.77	16.94
Ours	2.69	3.01	8.05	**1.57**	**1.73**	**4.71**	**4.68**	**5.01**	**13.56**	**2.16**	2.44	8.49	**2.06**	2.35	6.51	**2.84**	**3.36**	**11.54**

discontinuous point. Then we expand the detected boundaries by several pixels for covering the depth discontinuous regions. Evaluation of the depth discontinuous regions verifies the performance of the algorithm on the depth boundary areas.

1) Unharmed depth down-sampling: Fig.5 $\sim$ Fig.7 show the details of the 8 times up-sampling results of Art, $Laundry$ and $Dolls$, respectively. In which, results of the proposed method are also compared with the state-of-arts mentioned in Sec I-A. Table I and Table II present the 8 times BP and MSE evaluation results on all of the above 6 data sets, respectively. The proposed method shows great advantages in the evaluations of the depth discontinuous regions, which indicates that our method preserves the depth boundaries well. Comparisons show the superiorities of our method in both visual and quantitative evaluations.

We also evaluate the proposed method with different up-sampling rates of 4, 8, and 16, respectively. Table III and Table IV present the BP and MSE evaluation results on different regions and under different up-sampling rates, respectively. For better visualization, we provide the curve graphs in Fig.8 $\sim$ Fig.9. Results of different up-sampling rates verify the stability of the proposed method. Besides, it shows that our method performs quite well even under a large up-sampling rate.

2) Kinect-like depth down-sampling: As mentioned, depth missing exists in the Kinect-like depth maps. Up-sampling based on this kind of depth maps is more challenging. Fig.10 provides the 8 times up-sampling results based on the Kinect-like depth maps. Visual results demonstrate that our method effectively avoids the depth confusion artifacts around the disocclusion regions. That is, even in the case of depth-missing, the proposed method can recover the depth well and result in clear depth boundary.

The 8 times BP and MSE evaluations under this depth-missing situation are presented in Table V and Table VI, respectively. Quantitative evaluations also reinforce the good performance of the proposed method in the disocclusion regions.

B. Parameter Analysis

It's worth mentioning that the proposed method has good stability and parameter robustness. In our method, 4 parameters are used in the color image segmentation phase and 3 parameters are used in the interpolation phase.

In the SLIC superpixels, the initial number of superpixels k and the scale constant s is used. The roughly size of a superpixel can be calculated by N/k, where N denotes the total pixel number, that is to say, k is inversely proportional to the superpixel size. In other words, if k is small, the superpixel size will be large and this will result in a coarse segmentation, otherwise if k is large, a fine segmentation result will be

TABLE III
QUANTITATIVE UP-SAMPLING RESULTS FROM UNHARMED DEPTH DOWN-SAMPLING WITH UP-SAMPLING FACTORS OF 4, 8 AND 16 (IN BP). FOR EACH IMAGE, THE BEST PERFORMANCE IS SHOWN IN BOLD.

Methods	Laundry									Dolls								
	disocc.			*all*			*disc.*			*disocc.*			*all*			*disc.*		
	4	8	16	4	8	16	4	8	16	4	8	16	4	8	16	4	8	16
JBU	2.64	6.00	13.49	2.91	6.30	13.81	18.05	26.29	38.14	3.55	9.66	22.87	3.90	10.07	23.24	26.26	38.60	53.66
JBLM	1.95	4.56	10.88	2.21	4.84	11.19	13.34	20.28	31.51	2.89	8.56	19.69	3.23	8.97	20.07	20.66	32.57	43.12
MRF	4.32	7.29	14.94	4.60	7.58	15.25	31.22	34.85	45.16	3.76	8.96	22.29	4.11	9.35	22.65	33.02	41.73	54.68
JGU	3.07	5.71	12.43	3.36	6.00	12.75	21.53	28.33	35.20	3.83	8.53	20.72	4.29	9.02	21.12	29.01	40.33	59.12
TGV	6.54	7.37	15.65	6.77	7.65	15.94	28.00	32.04	40.09	3.56	6.20	19.96	3.94	6.65	20.09	26.03	38.10	52.47
AR	6.33	7.03	17.72	6.65	7.36	18.04	35.41	33.91	60.25	4.11	6.69	17.81	4.53	6.14	18.21	35.51	36.32	65.09
JAB	3.99	7.62	20.55	4.26	7.90	20.85	30.58	40.10	67.69	3.70	7.18	19.23	4.03	7.59	19.60	34.66	47.26	64.54
JTF	4.94	7.94	20.51	5.24	8.26	19.81	37.07	45.29	61.02	3.82	8.57	18.26	4.26	8.71	18.66	38.35	45.76	65.16
Ours	**1.31**	**2.63**	**8.39**	**1.55**	**2.91**	**8.71**	**10.72**	**16.48**	**27.84**	**1.60**	**4.74**	**13.56**	**1.99**	**5.22**	**14.03**	**15.62**	**25.09**	**40.02**

TABLE IV
QUANTITATIVE UP-SAMPLING RESULTS FROM UNHARMED DEPTH DOWN-SAMPLING WITH UP-SAMPLING FACTORS OF 4, 8 AND 16 (IN MSE). FOR EACH IMAGE, THE BEST PERFORMANCE IS SHOWN IN BOLD.

Methods	Laundry									Dolls								
	disocc.			*all*			*disc.*			*disocc.*			*all*			*disc.*		
	4	8	16	4	8	16	4	8	16	4	8	16	4	8	16	4	8	16
JBU	2.31	3.32	4.96	2.48	3.57	5.24	6.77	8.79	10.73	1.55	2.28	3.32	1.66	2.39	3.39	4.90	6.14	7.11
JBLM	2.49	3.58	5.41	2.67	3.85	5.71	7.32	9.61	11.66	1.66	2.41	3.52	1.79	2.55	3.60	5.31	6.66	7.79
MRF	**1.73**	**2.50**	4.13	**1.89**	**2.68**	4.41	**5.21**	**6.79**	9.27	**1.19**	1.72	2.83	**1.32**	1.82	2.90	**3.90**	4.74	6.21
JGU	2.23	2.98	4.39	2.46	3.29	4.76	6.75	8.43	9.86	1.56	2.21	2.99	1.71	2.37	3.11	5.10	6.23	7.04
TGV	2.62	3.08	3.81	2.74	3.48	4.36	6.76	8.53	9.92	1.43	1.85	5.30	1.55	1.96	5.35	4.47	5.25	8.31
AR	2.26	2.77	4.65	2.52	3.10	4.92	6.86	8.11	11.10	1.44	1.80	2.78	1.56	1.90	2.86	4.59	5.15	6.92
JAB	2.41	2.90	4.57	2.53	3.06	4.84	7.09	8.20	11.34	1.34	1.78	2.77	1.42	1.86	2.83	4.28	5.22	6.85
JTF	2.63	4.00	4.46	2.41	4.22	5.07	7.42	10.63	11.61	1.68	2.31	3.01	1.79	2.38	3.17	4.44	5.71	7.83
Ours	2.08	2.69	**3.42**	2.26	3.01	**3.82**	6.31	8.05	**8.76**	1.20	**1.57**	**2.33**	1.36	**1.73**	**2.46**	4.06	**4.71**	**5.74**

obtained. So that k is selected according to the expected superpixel size. Since the SLIC superpixels is used as a coarse segmentation in our method, the generated superpixels will be subdivided and recombined later, thus makes a good tolerance in the selection of k. However, too large k will result in time waste in the region splitting process and too small k leads to time waste in the region merging process, thus we set k around 800 in the experiments.

The scalar s is used to balance the proportion of the spatial distance and the intensity distance. When s is large, spatial proximity is more important, thus the resulting superpixels are more compact. When s is small, the resulting superpixels adhere to the image boundary more closely[30]. Generally, s is set in a range of $[1, 40]$, the smaller the more accurate the segmentation result, but the larger the time costs. For balancing the segmentation quality and the time costs, we set $s = 10$.

The splitting threshold th_s is used to split the incorrect-segmented regions after the SLIC. Thus the smaller th_s, the finer the segmentation result. However, too small th_s may result in lots of unnecessary over-segmented subregions. Since this parameter is mainly used for preserving the boundary of the up-sampled image, we provide its impact on the bad pixel rate of the depth discontinuous regions in Fig.11. The figure also presents the time costs along varies th_s, which shows a trade-off between the algorithm performance and the computational complexity.

Then a merging threshold th_m is used to limit the depth gap between the to be combined regions during the merging process, thus th_m should be relatively small. We set it no larger than 5 in the experiments. Fig.12 presents the magnified details under different th_s and th_m.

In the interpolation part, the two variances σ_c and σ_d are positively related to the up-sampling rate. We set $\sigma_c = 3\sqrt{r}/2$ and $\sigma_d = r$ empirically, where r denotes the up-sampling rate. The parameter γ in Eq.(20) will be impacted by different data sets, generally we set it between 0 and 0.1. Fig.13 provides the results of various γ on $Laundry$ data set.

C. Remark on Computational Cost

The computational cost is also evaluated, comparisons between the proposed method and the state-of-arts is shown in Table VII. As a simple bilateral filter based method, JBU process the target pixels in raster-scan order and spend almost constant time as long as the image resolutions are approximate. JBLM is resulted by adding a LM filter on the above joint bilateral filter, the processing time of the added

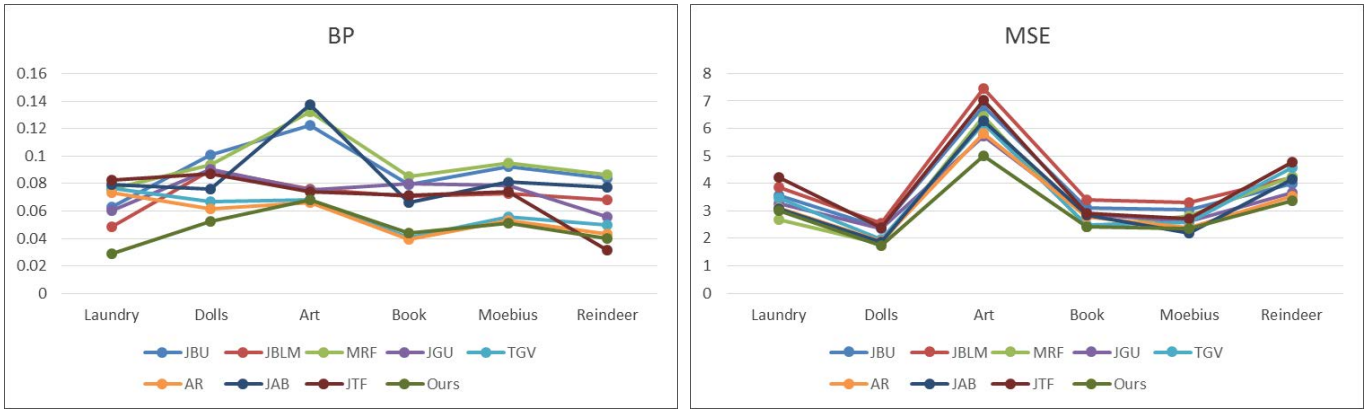


Fig. 8. 8 times up-sampling results of the 6 data sets (whole image) with different methods. Left: BP evaluations. Right: MSE evaluations.

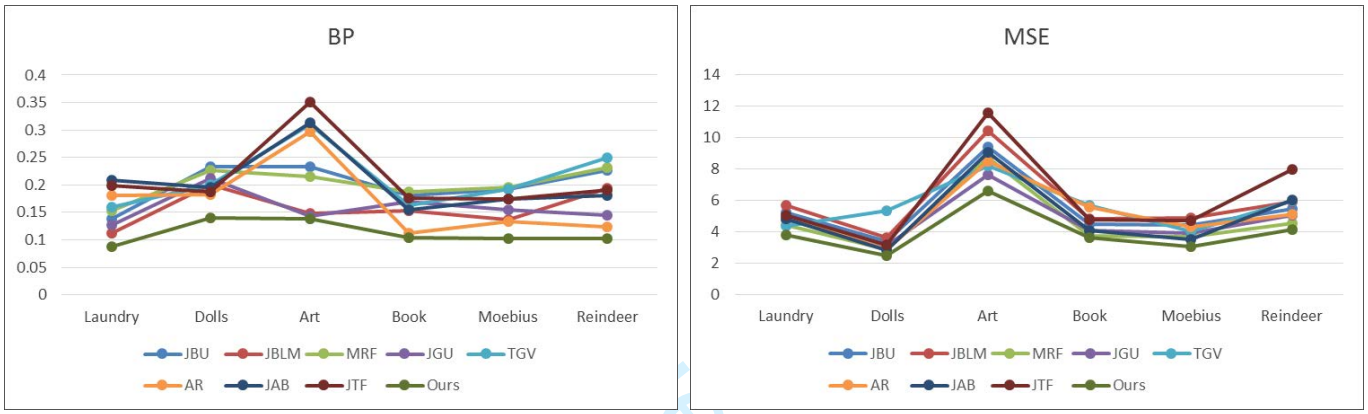
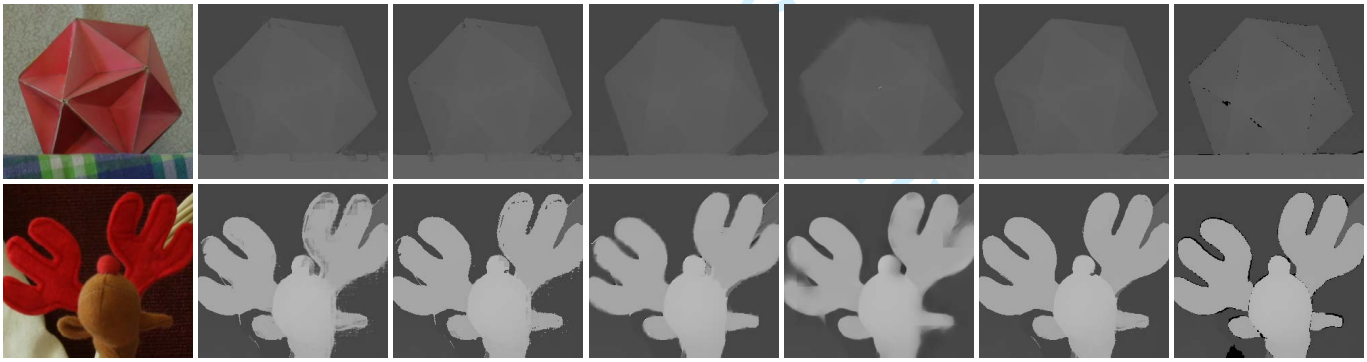


Fig. 9. 16 times up-sampling results of the 6 data sets (whole image) with different methods. Left: BP evaluations. Right: MSE evaluations..



(a) (b) (c) (d) (e) (f) (g)

Fig. 10. 8 times up-sampling results of $Moebius$ and $Reindeer$. (a) Color regions, and up-sampling results of (b) JBU, (c) JBLM, (d) TGV, (e) AR Model, (f) the proposed method, and (g) associated ground truth.

LM is positively related to the up-sampling rate. The MRF method and the AR Model method are two global methods and the running time is only related to the image resolution. The processing time of JGU is associated with both the image resolution and the up-sampling rate. TGVs time cost involves only the repeating times of the optimization, more further, the larger the up-sampling rate, the more repeating times are required. JAB and JTF are two filtering-based methods, both of their time costs increase with the enlarging of the up-sampling rate.

The proposed method includes the color image segmentation phase and the depth interpolation phase. The processing time of the first phase mainly depends on the image size and the splitting threshold th_s, and that of the second part is positively related to the up-sampling rate. We list the time costs of each phase in Table VII. Compared to existing methods, our method greatly reduces the computational complexity while maintaining high performance.

TABLE V
QUANTITATIVE UP-SAMPLING RESULTS FROM KINECT-LIKE DEPTH DOWN-SAMPLING WITH UP-SAMPLING FACTOR OF 8 (IN BP). FOR EACH IMAGE, THE BEST PERFORMANCE IS SHOWN IN BOLD.

Methods	Laundry			Dolls			Art			Books			Moebius			Reindeer		
	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.
JBU	5.67	5.99	25.42	9.24	9.71	37.35	11.17	11.68	38.91	7.71	8.16	39.73	8.88	9.37	45.01	8.09	8.45	38.41
JBLM	4.67	4.99	21.19	8.36	8.83	32.41	6.98	7.52	**20.53**	6.63	7.06	31.59	6.05	6.49	30.41	6.35	6.61	**24.15**
TGV	7.19	7.66	31.67	6.05	6.52	37.80	6.57	**6.87**	46.04	**3.38**	**3.88**	31.61	4.87	5.37	35.82	4.31	4.67	39.13
AR	10.69	11.92	74.13	10.43	10.13	73.59	10.66	11.67	82.62	7.49	8.01	53.84	8.92	9.48	51.50	7.15	7.61	55.20
Ours	**2.80**	**3.17**	**16.44**	**4.82**	**5.32**	**25.45**	**6.38**	6.91	25.10	4.33	4.78	**30.79**	**4.75**	**5.17**	**27.91**	**3.75**	**4.06**	24.88

TABLE VI
QUANTITATIVE UP-SAMPLING RESULTS FROM KINECT-LIKE DEPTH DOWN-SAMPLING WITH UP-SAMPLING FACTOR OF 8 (IN MSE). FOR EACH IMAGE, THE BEST PERFORMANCE IS SHOWN IN BOLD.

Methods	Laundry			Dolls			Art			Books			Moebius			Reindeer		
	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.	disocc.	all	disc.
JBU	3.28	4.11	10.24	2.14	2.27	5.91	6.56	6.96	17.78	2.93	3.14	10.12	2.72	2.89	7.55	3.51	3.87	12.71
JBLM	3.49	4.31	10.80	2.26	2.40	6.33	7.17	7.58	19.60	3.25	3.46	11.26	2.96	3.17	8.50	3.73	4.13	13.83
TGV	2.99	3.44	8.46	1.81	1.92	5.16	5.93	6.22	17.07	**2.19**	**2.41**	**8.53**	2.20	2.41	6.88	4.24	4.49	12.96
AR	2.74	3.13	**8.05**	1.82	1.92	5.14	5.61	5.86	15.70	2.61	2.82	9.41	2.10	2.27	6.36	3.12	3.42	11.77
Ours	**2.67**	**3.07**	8.11	**1.60**	**1.70**	**4.76**	**4.63**	**5.20**	**14.04**	2.39	2.64	8.95	**2.04**	**2.21**	**6.13**	**2.82**	**3.25**	**11.18**

TABLE VII
RUNNING TIME OF DIFFERENT METHODS WITH UP-SAMPLING FACTORS OF 4, 8 AND 16 (IN SECONDS). RUNNING TIMES OF THE COLOR IMAGE SEGMENTATION PART AND THE INTERPOLATION PART ARE LISTED IN THE LAST LINE.

Data sets	Dolls			Books		
Res. after up-sampling	1390×1110			1390×1110		
Up-sampling factors	4	8	16	4	8	16
Num. of target pixels	1446156	1518714	1536810	1446156	1518714	1536810
JBU	3.5	3.7	3.7	3.6	3.8	3.8
JBLM	22.8	37.5	65.6	15.1	22.3	38.0
MRF	553.0	542.1	544.9	553.2	546.4	546.8
JGU	325.4	467.5	631.1	326.5	484.3	670.6
TGV	1626.3	1636.1	2098.6	1612.3	1624.8	2112.3
AR	2566.5	2550.9	2545.2	2540.2	2535.5	2535.0
JAB	13.3	40.7	314.2	12.0	29.9	298.2
JTF	246.2	323.8	587.5	149.2	132.7	467.5
Ours	399.0 $(317.9 + 81.1)$	363.4 $(267.2 + 96.2)$	372.8 $(226.1 + 146.7)$	307.0 $(256.5 + 50.5)$	301.4 $(236.8 + 64.6)$	353.0 $(236.6 + 116.4)$

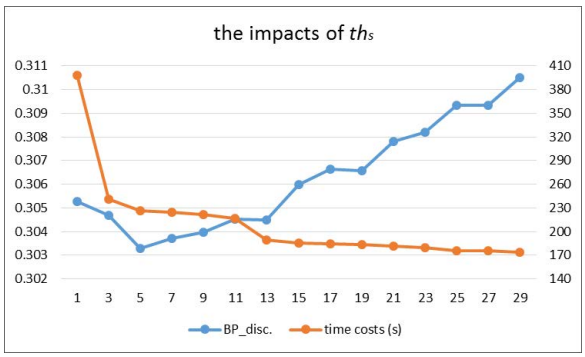


Fig. 11. The impacts of th_s on the bad pixel rate in the depth discontinuous regions ($BP\_disc.$) as well as the time costs.

V. CONCLUSIONS

We present a color image segmentation guided depth map up-sampling method in this paper. By definition, we segment the color image firstly, and use the segmented result to up-sample the depth maps. We carry out the SLIC superpixels, the region splitting and the region merging operations on the HR color image successively. Then we obtain the segmentation result that composed of certain numbers of homogeneous connected regions. This segmentation result further contributes to the region term, which together with the color term and the distance term, comprises the weighting factor of the proposed JTF. In the following JTF based interpolation, the targets are interpolated with the weighted averages of their neighboring seeds. The proposed method produces high quality up-sampled

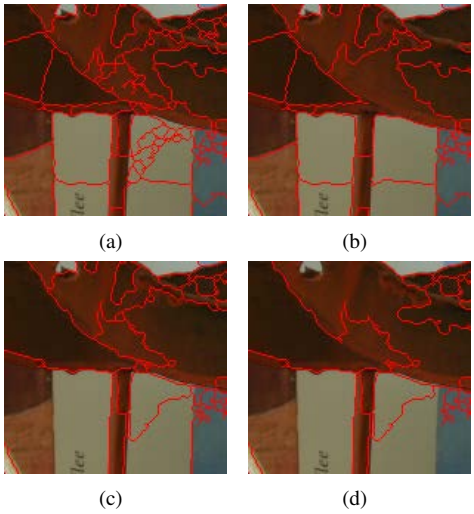


Fig. 12. Magnified splitting and merging results under varies th_s and th_m. Region splitting with (a) $th_s = 5$ and (b) $th_s = 25$, region merging with (c) $th_m = 5$ and (d) $th_m = 25$ under the condition that $th_s = 5$.

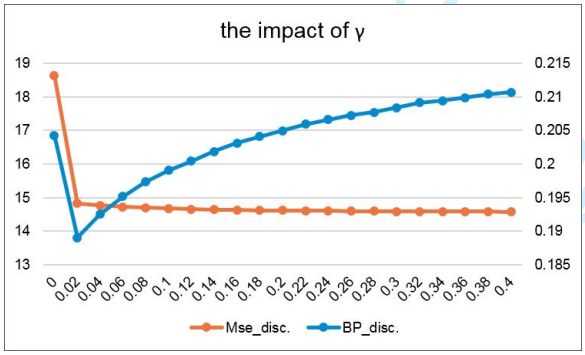


Fig. 13. The impacts of γ on BP and MSE in the depth discontinuous regions.

depth maps even under a relatively large up-sampling rate. Moreover, our method has a strong robustness, which makes the parameter settings flexible. Experimental results show that our method provides good performances in both subjective and objective evaluations.

REFERENCES

[1] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-d video," *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 453–465, June 2011.

[2] F. Shao, G. Jiang, M. Yu, K. Chen, and Y. S. Ho, "Asymmetric coding of multi-view video plus depth based 3-d video for view rendering," *IEEE Transactions on Multimedia*, vol. 14, no. 1, pp. 157–167, Feb 2012.

[3] B. Macchiavello, C. Dorea, E. M. Hung, G. Cheung, and W. T. Tan, "Loss-resilient coding of texture and depth for free-viewpoint video conferencing," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 711–725, April 2014.

[4] D. Ren, S. H. G. Chan, G. Cheung, V. Zhao, and P. Frossard, "Anchor view allocation for collaborative free viewpoint video streaming," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 307–322, March 2015.

[5] G. Petrazzuoli, T. Maugey, M. Cagnazzo, and B. Pesquet-Popescu, "Depth-based multiview distributed video coding," *IEEE Transactions on Multimedia*, vol. 16, no. 7, pp. 1834–1848, Nov 2014.

[6] T. Maugey, G. Petrazzuoli, P. Frossard, M. Cagnazzo, and B. Pesquet-Popescu, "Reference view selection in dibr-based multiview coding," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1808–1819, April 2016.

[7] K. Yoon and I. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 650–656, 2006.

[8] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (tof) cameras: A survey," *IEEE Sensors Journal*, vol. 11, no. 9, pp. 1917–1926, Sept 2011.

[9] K. Min-Koo, K. Dae-Young, and Y. Kuk-Jin, "Adaptive support of spatial-temporal neighbors for depth map sequence up-sampling," *IEEE Signal Processing Letters*, vol. 21, no. 2, 2014.

[10] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318–1334, Oct 2013.

[11] O. Choi and S. W. Jung, "A consensus-driven approach for structure and texture aware depth map upsampling," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3321–3335, Aug 2014.

[12] D. Chan, H. Buisman, C. Thebalt, , and S. Thrun, "A noise-aware filter for real-time depth upsampling," *in Workshop on M2SFA2, ECCV*, 2008.

[13] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3d-tof cameras," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 1623–1630.

[14] M. C. Yang and Y. C. F. Wang, "A self-learning approach to single image super-resolution," *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 498–508, April 2013.

[15] Z. Zhu, F. Guo, H. Yu, and C. Chen, "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec 2014.

[16] Z. Xiong, D. Xu, X. Sun, and F. Wu, "Example-based super-resolution with soft information and decision," *IEEE Transactions on Multimedia*, vol. 15, no. 6, pp. 1458–1465, Oct 2013.

[17] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transaction on Graphics*, vol. 26, no. 3, 2007.

[18] S. B. Lee, S. Kwon, and Y. S. Ho, "Discontinuity adaptive depth upsampling for 3d video acquisition," *Electronics Letters*, vol. 49, no. 25, pp. 1612–1614, December 2013.

[19] Y. S. Kang, S. B. Lee, and Y. S. Ho, "Depth map upsampling using depth local features," *Electronics Letters*, vol. 50, no. 3, pp. 170–171, 2014.

[20] P. F. Flezenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 24, no. 4, p. 603C619, 2002.

[21] M. Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," *CVPR*, 2013.

[22] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," *Proc. ICCV*, pp. 993–1000, 2013.

[23] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from rgb-d data using an adaptive autoregressive model," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3443–3458, Aug 2014.

[24] J. Kima, G. Jeonb, and J. C. Jeong, "Joint-adaptive bilateral depth map upsampling," *in Signal Processing ImageCommunication*, vol. 29, p. 506C513, 2014.

[25] K. H. Lo, Y. C. F. Wang, and K. L. Hua, "Edge-preserving depth map upsampling by joint trilateral filter," *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–14, 2017.

[26] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, Aug 2000.

[27] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: fast superpixels using geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, p. 2290, 2009.

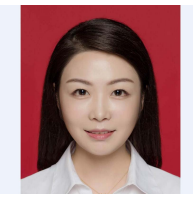
[28] Y. J. Liu, C. C. Yu, M. J. Yu, and Y. He, "Manifold slic: A fast method to compute content-sensitive superpixels," in *Computer Vision and Pattern Recognition*, 2016, pp. 651–659.

[29] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, p. 881C892, 2002.

[30] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, Nov 2012.

[31] B. Hill, T. Roger, and F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the cielab color-difference formula," *ACM Transactions on Graphics*, vol. 16, pp. 109–154, 1997.

[32] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, Jun 1994.
[33] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800–810, Aug 2001.
[34] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Transactions on Image Processing*, vol. 7, no. 12, pp. 1684–1699, Dec 1998.
[35] S. W. Jung, "Enhancement of image and depth map using adaptive joint trilateral filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 258–269, Feb 2013.
[36] K. H. Lo, Y. C. F. Wang, and K. L. Hua, "Joint trilateral filtering for depth map super-resolution," in *2013 Visual Communications and Image Processing (VCIP)*, Nov 2013, pp. 1–6.
[37] Http://vision.middlebury.edu/stereo/data/.
[38] S. M. Hong and Y. S. Ho, "Depth map refinement using superpixel label information," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec 2016, pp. 1–4.



Yiguo Qiao received the B.S. degree from Xidian University, Xian, China, in 2010. She is currently pursuing the Ph.D. degree at the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, Xian, China. Her research interests include free-viewpoint generation, depth map refinement, depth map up-sampling and some other 3DTV techniques.

Licheng Jiao (SM89) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Xian Jiaotong University, Xian, China, in 1984 and 1990, respectively. Currently, he is a Professor and Dean with the Electronic Engineering School, Xidian University, Xian, China. His research interests include neural networks, data mining, nonlinear intelligence signal processing, and communication.

Shuyuan Yang received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xian, China, in 2000, 2003, and 2005, respectively. She is currently a Professor with the School of Electronic Engineering, Xidian University. Her current research interests include compress sensing, machine learning, and intelligent information processing.

Biao Hou received the B.S. and M.S. degrees in mathematics from Northwest University, Xian, China, in 1996 and 1999, respectively, and the Ph.D. degree in circuits and systems from Xidian University, Xian, in 2003. Since 2003, he has been with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, where he is currently a Professor. His research interests include compressive sensing and Synthetic Aperture Radar image interpretation.